

IN

N₁MS ORF DNA Sequence [1 to 1116] -> 1-phase Translation

DNA sequence	1116 bp.	ATGGCTGCAACC ... AGTOGGAAGTGA	linear
1	1116	ATGGCTGCAACC ... AGTOGGAAGTGA	linear

complete new ETS ORF region DNA sequence (confirmed on 6/14/96).

37/ amino acids

$M_w = 41428 \text{ Dalton}$

1/1	31/11	61/21
ATG GCT GCA ACC TGT GAG ATT AGC AAC ATT TTT AGC AAC TAC TTC AGT GGG ATG TAC AGC TCG GAG GAC TCC ACC CTG GCC TCT GTT CCC Met ala ala thr cys glu ile ser asn ile phe ser asn tyr phe ser ala met tyr ser ser glu asp ser thr leu ala ser val pro	91/31	151/51
CTT GCT GGC ACC TTT GGG GGC GAT GAC TTG GTA CTG ACC CTG AGC AAC CCC CAG ATG TCA TTG GAG GGT ACA GAG AAG GGC AGC TCG TTG pro ala ala thr phe gly ala asp asp leu val leu thr leu ser asn pro gln met ser leu glu gly thr glu lys ala ser trp leu	211/71	241/81
181/61	211/71	331/111
GGG GAA CAG CCC CAG TTC TGG TCG AAG ACG CAG GTT CTG GAC TGG ATC AGC TAC CAA GTG GAG AAG AAC TAC GAC GCA AGC GGC ATT gly glu gly gln pro gln phe trp ser lys thr gln val leu asp trp ile ser tyr gln val glu lys asn lys tyr asp ala ser ala ile	301/101	421/141
271/91	301/101	511/171
GAG TTC TCA CGA TGT GAC ATG GAT GGC GCC ACC CTC TCC AAC TGT GCT GGC CTG GAG CTG ATT GAG CTG CTG GAG AAG GAT GGC ATG GGC TTC asp phe ser arg cys asp met asp gly ala thr leu cys asn cys ala leu glu glu leu arg leu val phe gly pro leu gly asp gln	391/131	601/201
361/121	391/131	691/231
CTC CAT GGC CTC GCA GAC CTC ACT TCC ACC TCT TCT GAT GAG CTC AGT TCG ATC ATT GAG CTG CTG GAG AAG GAT GGC ATG GGC TTC leu his ala gln leu arg asp leu thr ser ser asp glu leu ser trp ile ile glu leu leu glu lys asp gly met ala phe	481/161	871/291
451/151	481/161	961/321
CAG GAG GGC CTA GAC CCA GGG CCC TTT GAC CAG GGC AGC CCC TTT GCC CAG GAG CTG CTG GAG GAC GGT CAG CAA GCC AGC CCC TAC CAC gln glu ala leu asp pro gly pro phe asp gln gly ser pro phe ala gln glu leu leu asp asp gly gln gln ala ser pro tyr his	571/191	1051/351
541/181	571/191	1101/361
CGC GGC AGC TGT GGC CCA GGA GGC CCC TCC CTT GGC AGC TCT GAC CTC TCC ACC GCA GGG ACT GGT GCT TCT CCG AGC TCC CAC TCC TCA pro gly ser cys gly ala gly ala pro ser pro gly ser ser asp val ser thr ala gly thr gly ala ser arg ser ser his ser ser	661/221	1201/341
631/211	661/221	1301/381
GAC TCC GGT GCA AGT GAC GTG GAC CTG GAT CCC ACT GAT GGC MAG CTC TTC CCC AGC GAT GGT TTT GGT GAC TGC MAG AAG GGG GAT CCC asp ser gly gly ser asp val asp leu asp pro thr asp gly lys leu phe pro ser asp gly phe arg asp cys lys lys gly asp pro	751/251	1401/421
721/241	751/251	1501/461
AAG CAC CAG CCG AAG CCG AAA CGA GGC CCG CCC CGA AAG CTG AGC AAA GAG TAC TGG GAC TGT CTG GAG GGC AAG AGC AAG CAC GCG CCC gly his his gly lys arg lys arg pro arg lys leu ser lys glu tyr trp asp cys leu glu gly lys ser lys his ala pro	841/281	1601/501
811/271	841/281	1701/541
AGA GGC ACC CAC CTG TGG GAG TTC ATC CCG GAT CTA CTC ATC CAC CCG GAG CTC AAC GAG GGC CTC ATG AAG TCG GAG AAT CCG CAT GAA arg gly thr his leu trp glu phe ile arg asp ile leu ile his pro glu leu asn glu gly leu met lys trp glu asn arg his glu	931/311	1801/581
901/301	931/311	1901/621
GCG GTC TTC aag TTC CTG GGC TCC GAG GCT GTG GCC CAA CTA TGG GGC CAA MAG AAA MAG AAC AGC AAC ATG ACC TAC GAG AAG CTG AGC gly val thr lys phe leu arg ser glu ala val ala gln leu trp gly gln lys lys asn ser asn met thr tyr glu lys leu ser	1021/341	2001/661
991/331	1021/341	2101/701
CGC GGC ATT AGG TAC TAC AAA CCG GAG ATC CTG GAA CCG GTG GAT GGC CGA CTC CCG CGA CTC GTC TAC AAG TTT GGC AAA AAC TCA AGC GGC arg ala met arg tyr tyr cys arg gly leu glu arg val asp gly arg arg leu val tyr lys phe gly lys asn ser ser gly	1101/371	2201/741
1081/361	1101/371	2301/781
TCG AAG GAG GAA GAG GTT CTC CAG AGT CCG AAC TGA trp lys glu glu val leu gln ser arg asn opa		

Fig. 1

Fig. 2a

MAATCEISNIFSNYFSAMYSSDSTLASVPPAATFGADDLVLTLSNPQMSLEG	53
TEKASWLGEQ	106
<u>PQFWSKTQVLDWISYQVEKNKYDASAIDFSRCDMDGATLCNCA</u>	
LEELRLVFGPLGDQLHAQLRDLTSSSSDELSTWIIELLEKDGMAFQEAALDPGPF	159
DQGSPPAQELLDDGQQASPYHPGSCGAG	212
<u>APSPGSSDVSTAGTGASRSSHSSDS</u>	
<u>GGSDVDLDPTDGKLLFPDGFDRDCKKG</u>	265
SKHAPRGTHLWEEFIRDILLHPELNEGLMKWENRHEGVFKFLRSEAVAQLWGQK	318
<u>KKNSNMTYEKLSRAMRYYYKREILERVDGRRLLVYKF</u>	371
<u>GKNSSGWKEEEVLQSRN</u>	

Fig. 2b

Consensus⁺: P W V W W E G LC
 ESX (64-103): PQFWSKTQVLDWISYQVEKNKYDASAIDEFRCMDGATLC
 P+ W++T V DW+ + V N++ +DF + M+GA LC
 ETS-1 (69-106): PRQWTEHVVDWMWAV--NEFSLKGVDFQKFCMNGAALC

Fig. 2c

ESX (188-238): APSPGSSDVSTAGTGASRSSHSSDGGSDVDLDPTDGKLFPSDGFRCCKG
 APS S S++ + S SS SS S SD + + L PS F G
 SOX4 (370-420): APSSAPSHASSASSHSSSSSSGSSDDEFEDDLLDNPSSNFESMSLG

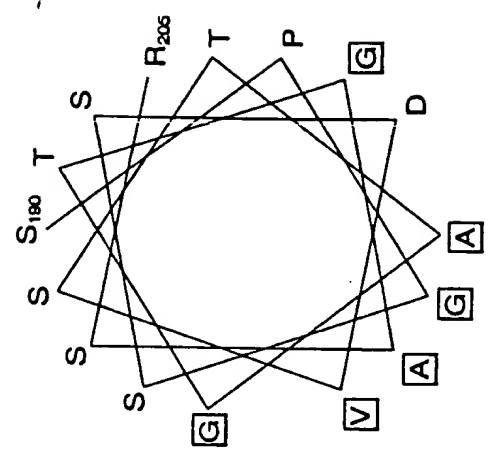


Fig. 2d

Consensus⁺: α_1 β_1 β_2 α_2 "turn" α_3 β_3 "wing" β_4
 LWQFLL LL D I W FK VAR WG K P MNY KLSR LRYYY I K G R Y F
 * * * * *
 ESX (274-354): HLWEEFIRDILIHPELNEGLMKWENRHEGVFKFLRSEAVAQLWGQKKNSNMTYEKLSRAMRYYYKREILERVDRRLVYKF
 +LWEEF+ +L +KW R +G+FK + S+AV++LWG+ K +M YE + RA+RYYY+R IL +V+G+RLVY+F
 ELF-1 (209-289): YLWEEFLLALQDKATCPKYIKWTQREKGIFKLVDKAVSRVLWGKHKNKPDNMNYETMGRALRYYYQRGILAKVEGQRLVYQF

Human ESX Protein Sequence

MAATCEISNIFSNYFSAMYSSDSTLASVPPAATFGADDLVLTLSNPQMSLEG 53
 TEKASWLGEQPFWSKTQVLDWISYQVEKNKYDASAIQFSRCDMDGATLCNCA 106
 LEELRLVFGPLGDLHAQLRDLTSSSSDELSWIELLEGGMAFQALDPCPF 159
 DQGSPPAQELLDDGQQAAPYHPGSCGAGAPSGSSDVSTAGTGASRSSHSDS 212
 GGSDVDLDPDTDGKLFPSDGFRCCKKDPKHGKRKRGRPRKLSKEYWOCLEGKK 265
 SKHAPRGTHLWEFIRDILIHPELNEGLMKWENRHEGVFKFLRSEAVAQLWGQK 318
 KKNSNMTYEKLSRAMRYYYKREILERVODGRRLLVYKFGKNSSGWKEEVLQSRN 371

Fig. 2e

mESX Genomic Organization and Gene Product

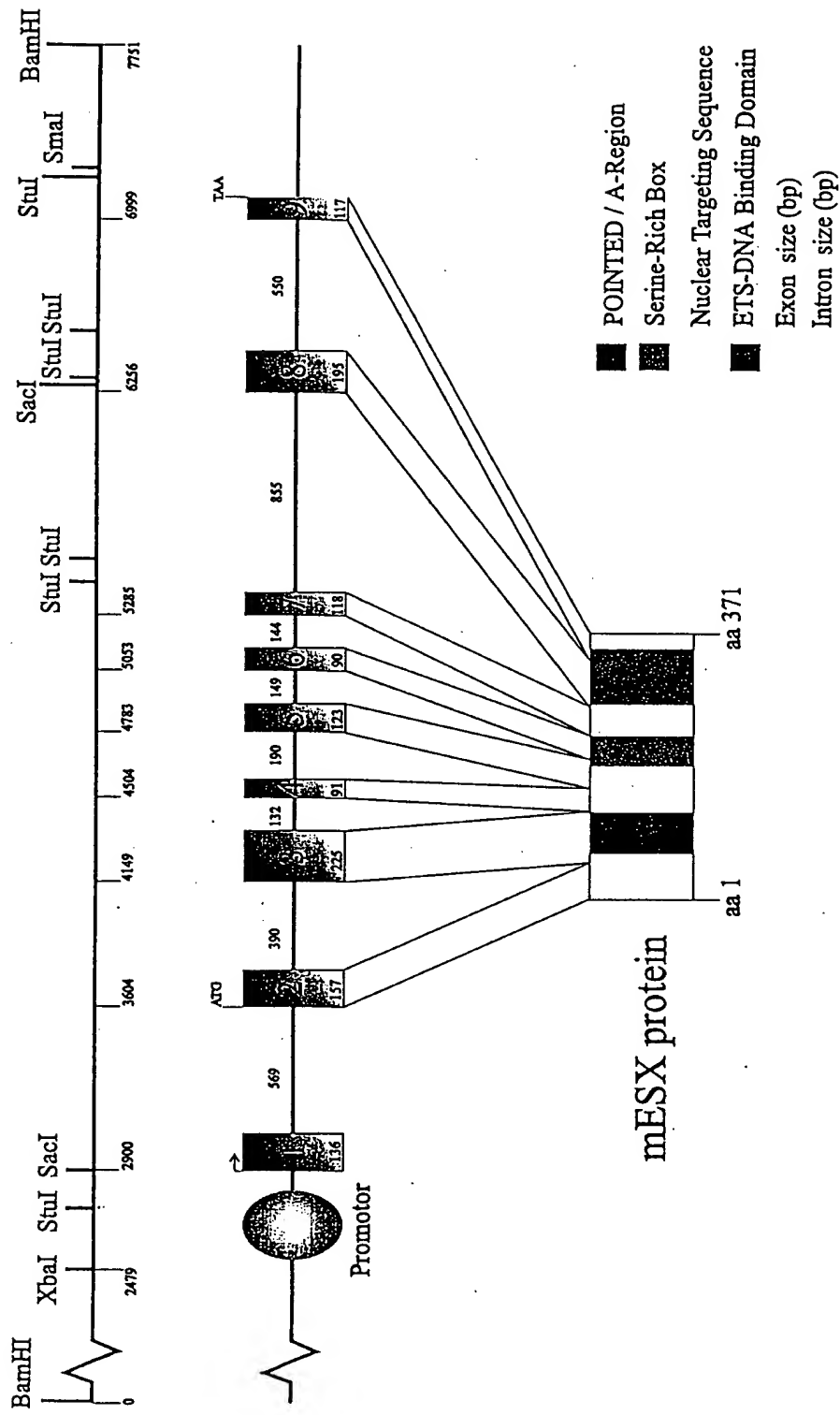


Fig. 3

Human ESX Exon/Intron Junctions

1/1	ATG GCT GCA ACC TGT GAG ATT AGC AAC ATT	31/11	TTT AGC AAC TAC TTC AGT GCG ATG TAC AGC
Met ala ala thr cys glu ile ser asn ile		phe ser asn tyr phe ser ala met tyr ser	
61/21	TCG GAG GAC TCC ACC CTG GCC TCT GTT CCC	91/31	CCT GCT GCC ACC TTT GGG GCC GAT GAC TTG
ser glu asp ser thr leu ala ser val pro		pro ala ala thr phe gly ala asp asp leu	
121/41	GTA CTG ACC CTG AGC AAC CCC CAG ATG TCA	151/51	TTG GAG GGT ACA GAG AAG GCC AGC TGG TTG
val leu thr leu ser asn pro gln met ser		leu glu gly thr glu lys ala ser trp leu	
181/61	GGG GAA CAG CCC CAG TTC TGG TCG AAG ACG	211/71	CAG GTT CTG GAC TGG ATC AGC TAC CAA GTG
gly glu gln pro gln phe trp ser lys thr		gln val leu asp trp ile ser tyr gln val	
241/81	GAG AAG AAC AAG TAC GAC GCA AGC GCC ATT	271/91	GAC TTC TCA CGA TGT GAC ATG GAT GGC GCC
glu lys asn lys tyr asp ala ser ala ile		asp phe ser arg cys asp met asp gly ala	
301/101	ACC CTC TGC AAT TGT GCC CTT GAG GAG CTG	331/111	CGT CTG GTC TTT GGG CCT CTG GGG GAC CAA
thr leu cys asn cys ala leu glu glu leu		arg leu val phe gly pro leu gly asp gln	
361/121	CTC CAT GCC CAG CTG CGA GAC CTC ACT TCC	391/131	AGC TCT TCT GAT GAG CTC AGT TGG ATC ATT
leu his ala gln leu arg asp leu thr ser	E3▼E4	ser ser ser asp glu leu ser trp ile ile	
421/141	GAG CTG CTG GAG AAG GAT GGC ATG GCC TTC	451/151	CAG GAG GCC CTA GAC CCA GGG CCC TTT GAC
glu leu leu glu lys asp gly met ala phe		gln glu ala leu asp pro gly pro phe asp	E4▼E5
481/161	CAG GGC AGC CCC TTT GCC CAG GAG CTG CTG	511/171	GAC GAC GGT CAG CAA GCC AGC CCC TAC CAC
gln gly ser pro phe ala gln glu leu leu		asp asp gly gln gln ala ser pro tyr his	
541/181	CCC GGC AGC TGT GGC GCA GGA GCC CCC TCC	571/191	CCT GGC AGC TCT GAC GTC TCC ACC GCA GGC
pro gly ser cys gly ala gly ala pro ser		pro gly ser ser asp val ser thr ala gly	E5▼E6
601/201	ACT GGT GCT TCT CGG AGC TCC CAC TCC TCA	631/211	GAC TCC GGT GGA AGT GAC GTG GAC CTG GAT
thr gly ala ser arg ser ser his ser ser		asp ser gly gly ser asp val asp leu asp	
661/221	CCC ACT GAT GGC AAG CTC TTC CCC AGC GAT	691/231	GGT TTT CGT GAC TGC AAG AAG GGG GAT CCC
pro thr asp gly lys leu phe pro ser asp	E6▼E7	gly phe arg asp cys lys lys gly asp pro	
721/241	AAG CAC GGC AAG CGG AAA CGA GGC CGG CCC	751/251	CGA AAG CTG AGC AAA GAG TAC TGG GAC TGT
lys his gly lys arg lys arg gly arg pro		arg lys leu ser lys glu tyr trp asp cys	
781/261	CTC GAG GGC AAG AAG AGC AAG CAC GCG CCC	811/271	AGA GGC ACC CAC CTG TGG GAG TTC ATC CCG
leu glu gly lys lys ser lys his ala pro	E7▼E8	arg gly thr his leu trp glu phe ile arg	
841/281	GAC ATC CTC ATC CAC CCG GAG CTC AAC GAG	871/291	GGC CTC ATG AAG TGG GAG AAT CGG CAT GAA
asp ile leu ile his pro glu leu asn glu		gly leu met lys trp glu asn arg his glu	
901/301	GGC GTC TTC aag TTC CTG CGC TCC GAG GCT	931/311	GTG GCC CAA CTA TGG GGC CAA AAG AAA AAG
gly val phe lys phe leu arg ser glu ala		val ala gln leu trp gly gln lys lys lys	
961/321	AAC AGC AAC ATG ACC TAC GAG AAG CTG AGC	991/331	CGG GCC ATG AGG TAC TAC TAC AAA CGG GAG
asn ser asn met thr tyr glu lys leu ser	E8▼E9	arg ala met arg tyr tyr tyr lys arg glu	
1021/341	ATC CTG GAA CGG GTG GAT GGC CGG CGA CTC	1051/351	GTC TAC AAG TTT GGC AAA AAC TCA AGC GGC
ile leu glu arg val asp gly arg arg leu		val tyr lys phe gly lys asn ser ser gly	
1081/361	TGG AAG GAG GAA GAG GTT CTC CAG AGT CGG	1111/371	AAC TGA
trp lys glu glu glu val leu gln ser arg		asn OPA	

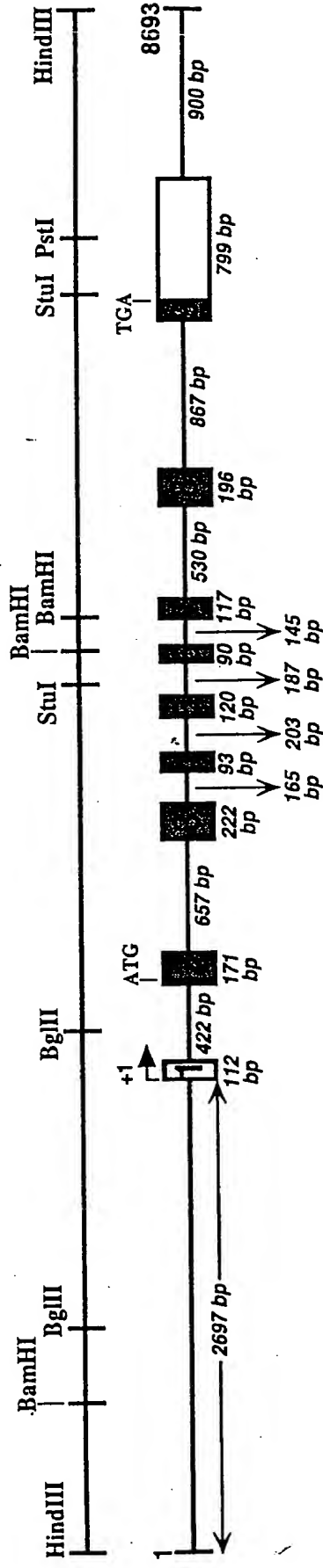
Fig. 4

1. **Introduction**
 2. **Background**
 3. **Methodology**
 4. **Results**
 5. **Discussion**
 6. **Conclusion**
 7. **References**
 8. **Appendix**
 9. **Figure 1**
 10. **Figure 2**
 11. **Figure 3**
 12. **Figure 4**
 13. **Figure 5**
 14. **Figure 6**
 15. **Figure 7**
 16. **Figure 8**
 17. **Figure 9**
 18. **Figure 10**
 19. **Figure 11**
 20. **Figure 12**
 21. **Figure 13**
 22. **Figure 14**
 23. **Figure 15**
 24. **Figure 16**
 25. **Figure 17**
 26. **Figure 18**
 27. **Figure 19**
 28. **Figure 20**
 29. **Figure 21**
 30. **Figure 22**
 31. **Figure 23**
 32. **Figure 24**
 33. **Figure 25**
 34. **Figure 26**
 35. **Figure 27**
 36. **Figure 28**
 37. **Figure 29**
 38. **Figure 30**
 39. **Figure 31**
 40. **Figure 32**
 41. **Figure 33**
 42. **Figure 34**
 43. **Figure 35**
 44. **Figure 36**
 45. **Figure 37**
 46. **Figure 38**
 47. **Figure 39**
 48. **Figure 40**
 49. **Figure 41**
 50. **Figure 42**
 51. **Figure 43**
 52. **Figure 44**
 53. **Figure 45**
 54. **Figure 46**
 55. **Figure 47**
 56. **Figure 48**
 57. **Figure 49**
 58. **Figure 50**
 59. **Figure 51**
 60. **Figure 52**
 61. **Figure 53**
 62. **Figure 54**
 63. **Figure 55**
 64. **Figure 56**
 65. **Figure 57**
 66. **Figure 58**
 67. **Figure 59**
 68. **Figure 60**
 69. **Figure 61**
 70. **Figure 62**
 71. **Figure 63**
 72. **Figure 64**
 73. **Figure 65**
 74. **Figure 66**
 75. **Figure 67**
 76. **Figure 68**
 77. **Figure 69**
 78. **Figure 70**
 79. **Figure 71**
 80. **Figure 72**
 81. **Figure 73**
 82. **Figure 74**
 83. **Figure 75**
 84. **Figure 76**
 85. **Figure 77**
 86. **Figure 78**
 87. **Figure 79**
 88. **Figure 80**
 89. **Figure 81**
 90. **Figure 82**
 91. **Figure 83**
 92. **Figure 84**
 93. **Figure 85**
 94. **Figure 86**
 95. **Figure 87**
 96. **Figure 88**
 97. **Figure 89**
 98. **Figure 90**
 99. **Figure 91**
 100. **Figure 92**
 101. **Figure 93**
 102. **Figure 94**
 103. **Figure 95**
 104. **Figure 96**
 105. **Figure 97**
 106. **Figure 98**
 107. **Figure 99**
 108. **Figure 100**
 109. **Figure 101**
 110. **Figure 102**
 111. **Figure 103**
 112. **Figure 104**
 113. **Figure 105**
 114. **Figure 106**
 115. **Figure 107**
 116. **Figure 108**
 117. **Figure 109**
 118. **Figure 110**
 119. **Figure 111**
 120. **Figure 112**
 121. **Figure 113**
 122. **Figure 114**
 123. **Figure 115**
 124. **Figure 116**
 125. **Figure 117**
 126. **Figure 118**
 127. **Figure 119**
 128. **Figure 120**
 129. **Figure 121**
 130. **Figure 122**
 131. **Figure 123**
 132. **Figure 124**
 133. **Figure 125**
 134. **Figure 126**
 135. **Figure 127**
 136. **Figure 128**
 137. **Figure 129**
 138. **Figure 130**
 139. **Figure 131**
 140. **Figure 132**
 141. **Figure 133**
 142. **Figure 134**
 143. **Figure 135**
 144. **Figure 136**
 145. **Figure 137**
 146. **Figure 138**
 147. **Figure 139**
 148. **Figure 140**
 149. **Figure 141**
 150. **Figure 142**
 151. **Figure 143**
 152. **Figure 144**
 153. **Figure 145**
 154. **Figure 146**
 155. **Figure 147**
 156. **Figure 148**
 157. **Figure 149**
 158. **Figure 150**
 159. **Figure 151**
 160. **Figure 152**
 161. **Figure 153**
 162. **Figure 154**
 163. **Figure 155**
 164. **Figure 156**
 165. **Figure 157**
 166. **Figure 158**
 167. **Figure 159**
 168. **Figure 160**
 169. **Figure 161**
 170. **Figure 162**
 171. **Figure 163**
 172. **Figure 164**
 173. **Figure 165**
 174. **Figure 166**
 175. **Figure 167**
 176. **Figure 168**
 177. **Figure 169**
 178. **Figure 170**
 179. **Figure 171**
 180. **Figure 172**
 181. **Figure 173**
 182. **Figure 174**
 183. **Figure 175**
 184. **Figure 176**
 185. **Figure 177**
 186. **Figure 178**
 187. **Figure 179**
 188. **Figure 180**
 189. **Figure 181**
 190. **Figure 182**
 191. **Figure 183**
 192. **Figure 184**
 193. **Figure 185**
 194. **Figure 186**
 195. **Figure 187**
 196. **Figure 188**
 197. **Figure 189**
 198. **Figure 190**
 199. **Figure 191**
 200. **Figure 192**
 201. **Figure 193**
 202. **Figure 194**
 203. **Figure 195**
 204. **Figure 196**
 205. **Figure 197**
 206. **Figure 198**
 207. **Figure 199**
 208. **Figure 200**
 209. **Figure 201**
 210. **Figure 202**
 211. **Figure 203**
 212. **Figure 204**
 213. **Figure 205**
 214. **Figure 206**
 215. **Figure 207**
 216. **Figure 208**
 217. **Figure 209**

■	Ets	■	USF
■	AP-2	■	Oct
■	extended palindrome	■	NF-κB
■	SP1/GC box	□ □	CCAAT box & Inr element

1

Human ESX Genomic DNA Structure



Mouse ESX Genomic DNA Structure

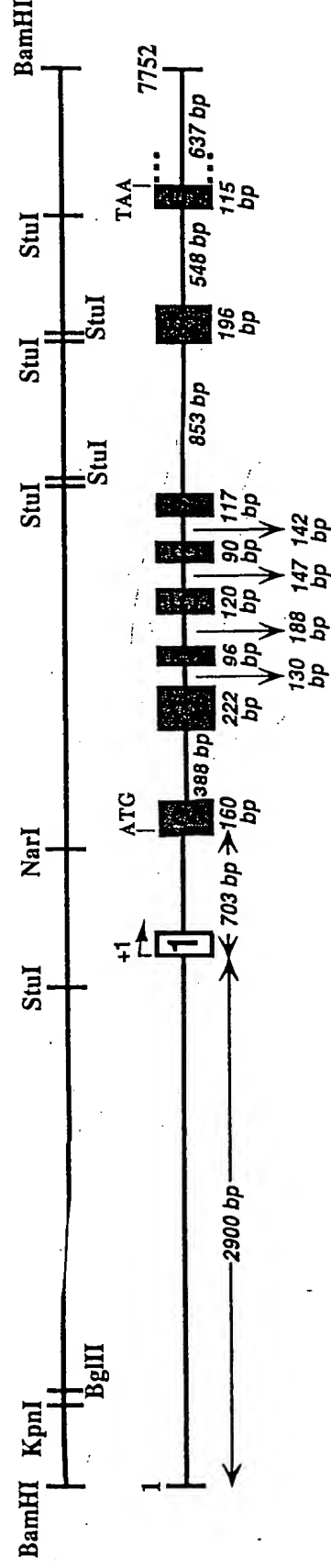
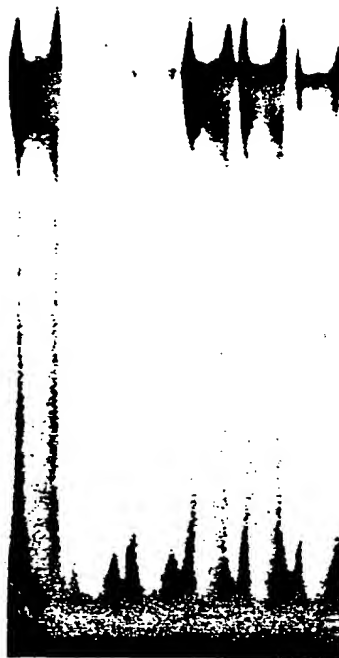


Fig. 7

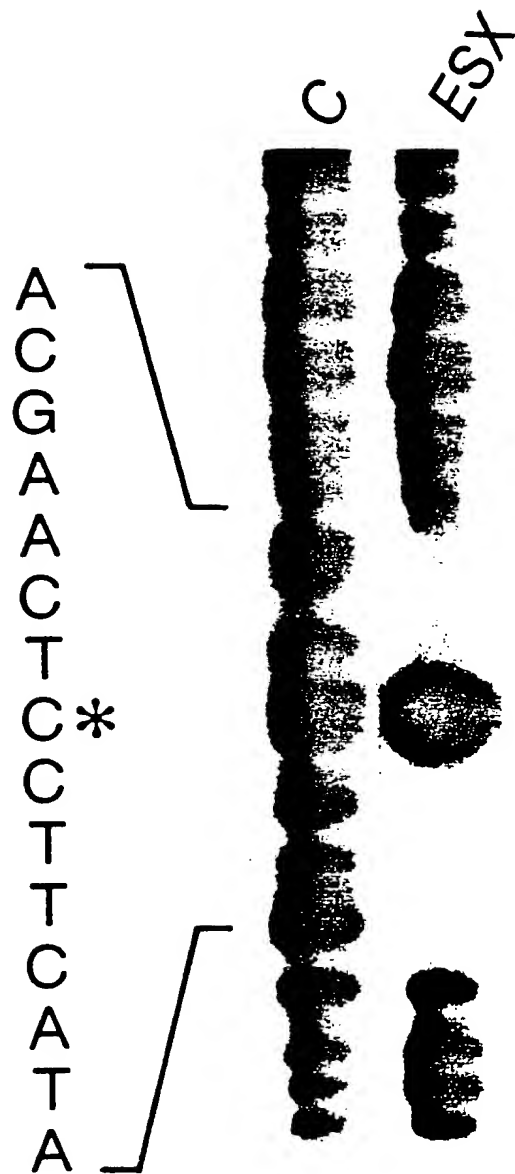
Fig. 8a

m1 m2 m3 m4 m5



WT 5' GGAGGAGGGCTGCTTGAGGAAGTATAAGAAT 3'
m1 5' ----- TA ----- 3'
m2 5' ----- C ----- 3'
m3 5' ----- AG ----- 3'
m4 5' ----- CC ----- 3'
m5 5' ----- C ----- 3'

Fig. 8b



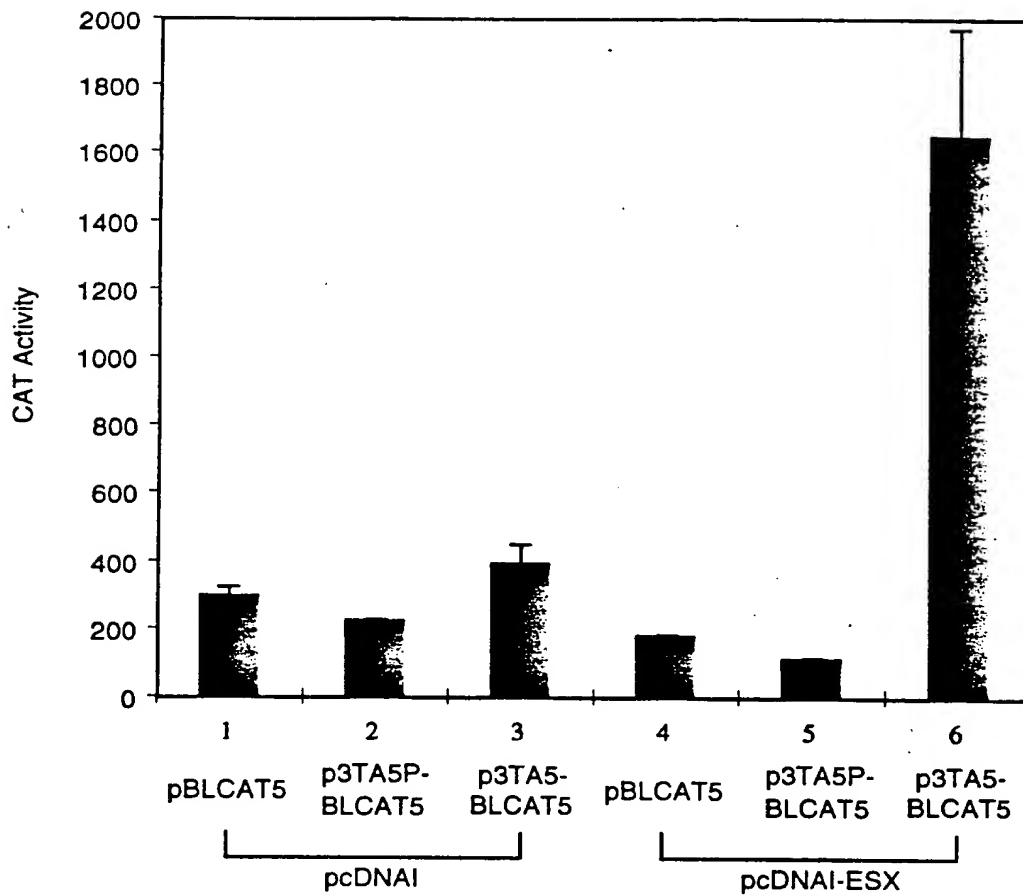


Fig. 8c

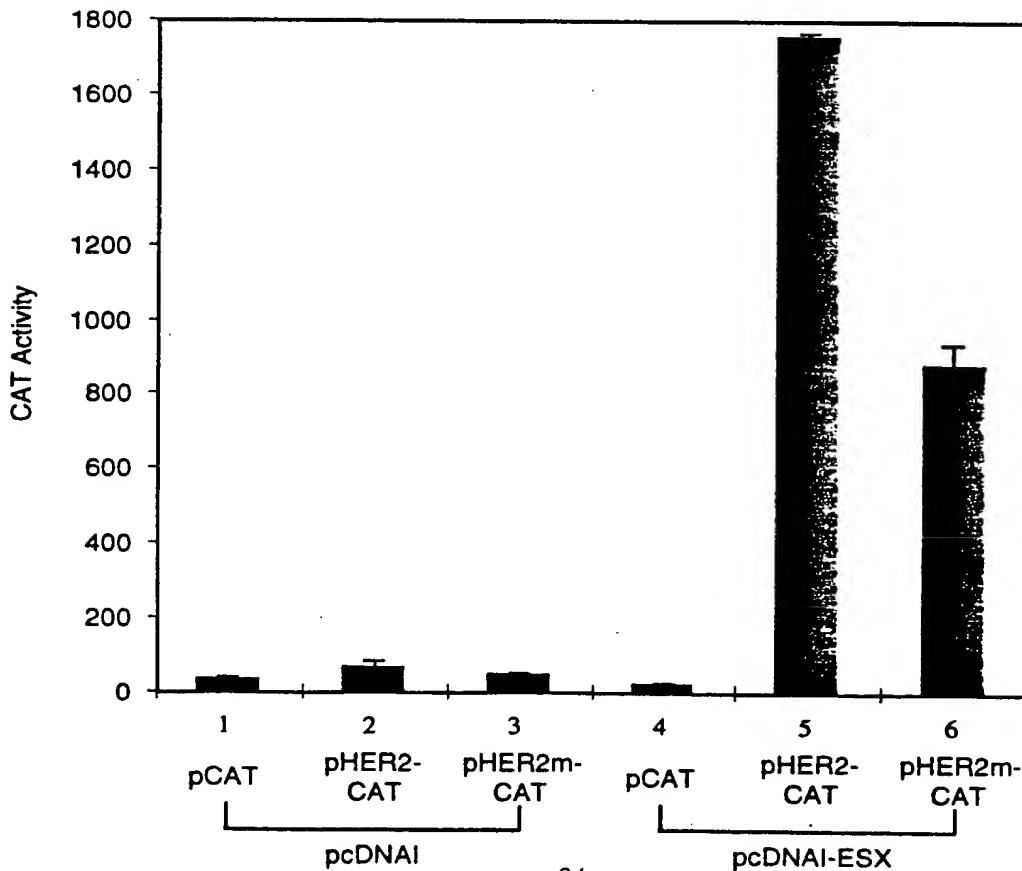
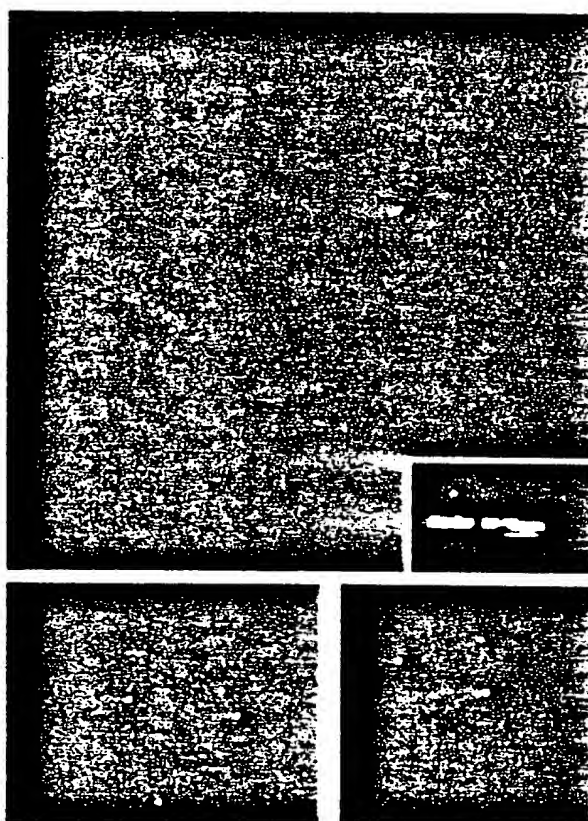


Fig. 8d

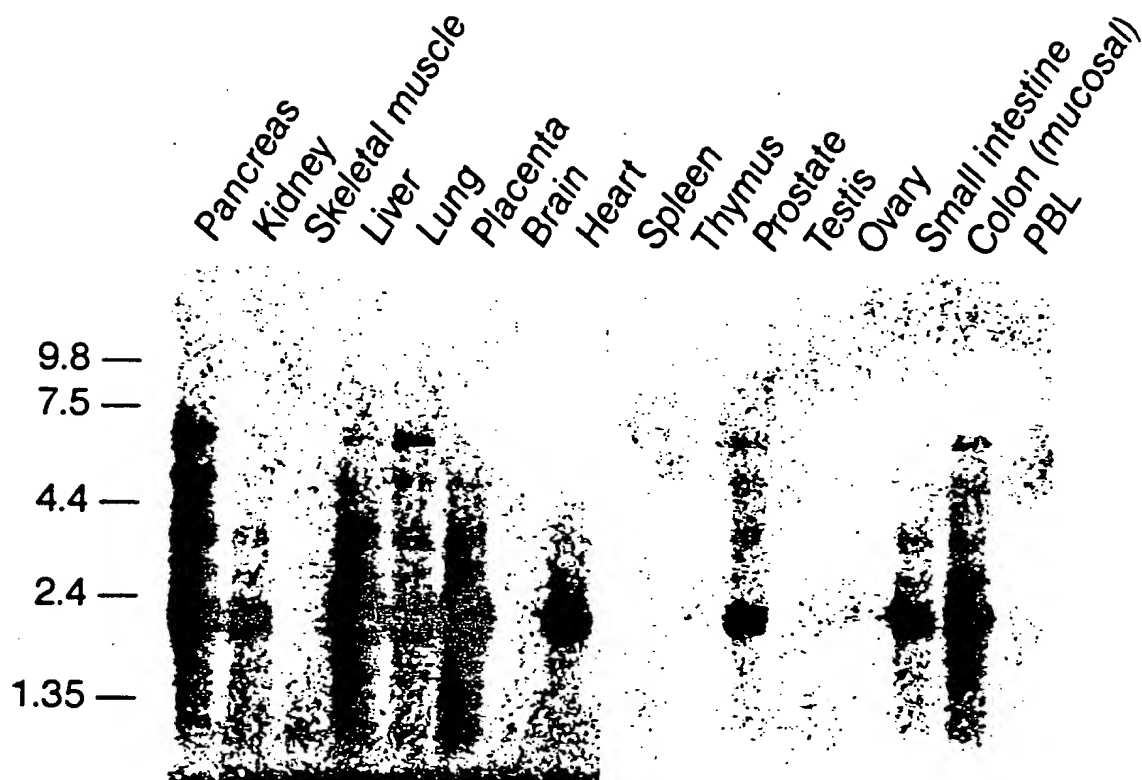
Fig. 8e



08978217 112597

08978217.112597

Fig. 9a



HMEC
 HBL-100
 MCF-10A
 MCF-7
 MDA-231
 BT-474
 SK-BR-3
 ZR-75-1

4.1 kb —
 2.2 kb —



HER2- HER2+

Fig. 9b

08978217 112597

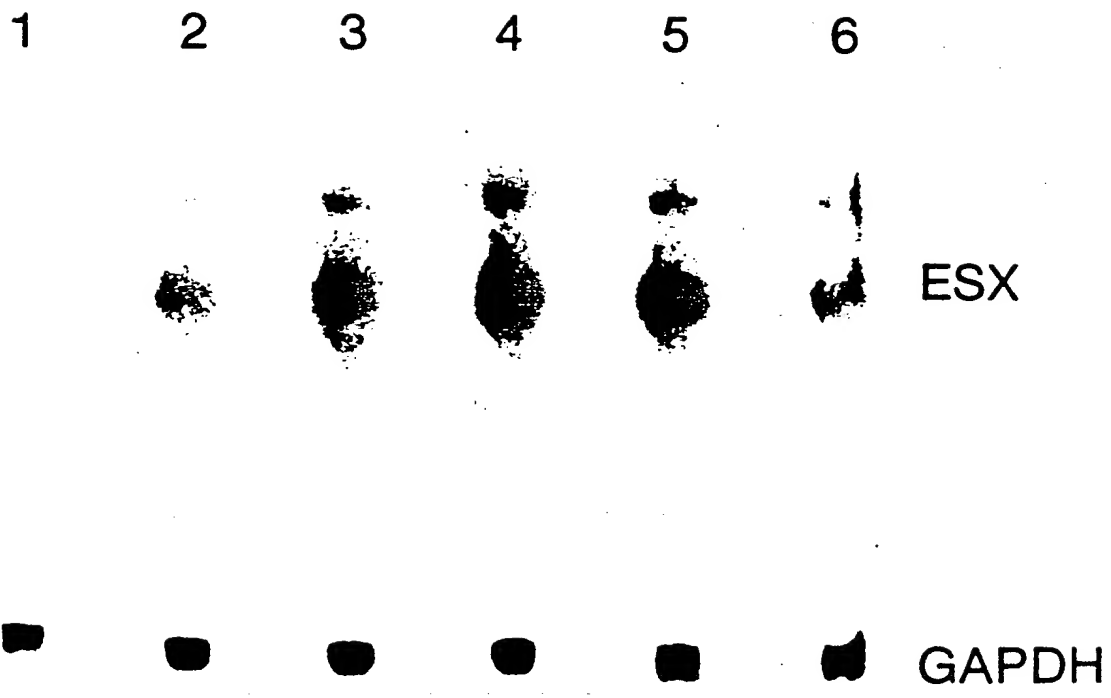


Fig. 9c

08978217 112597

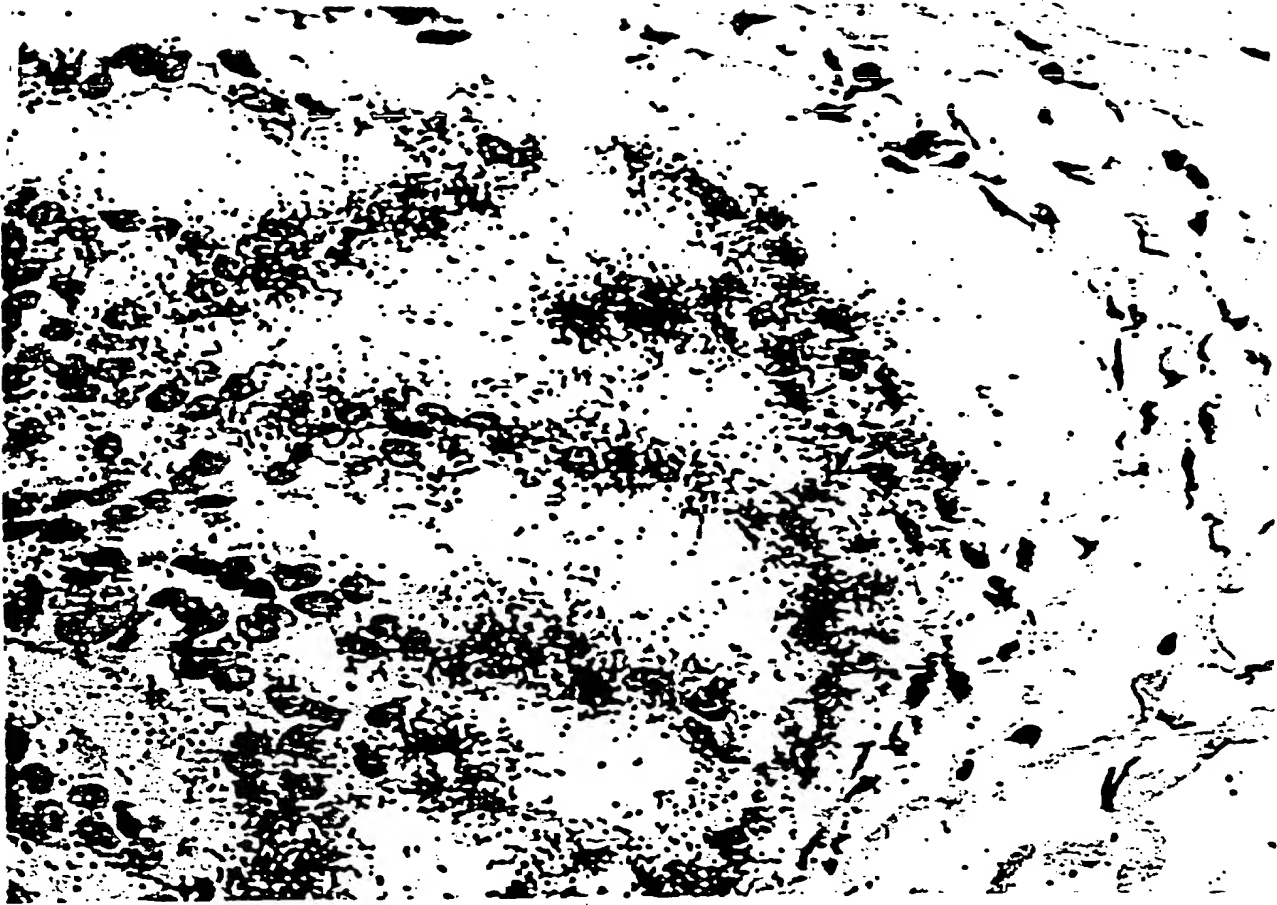


Fig. 10a

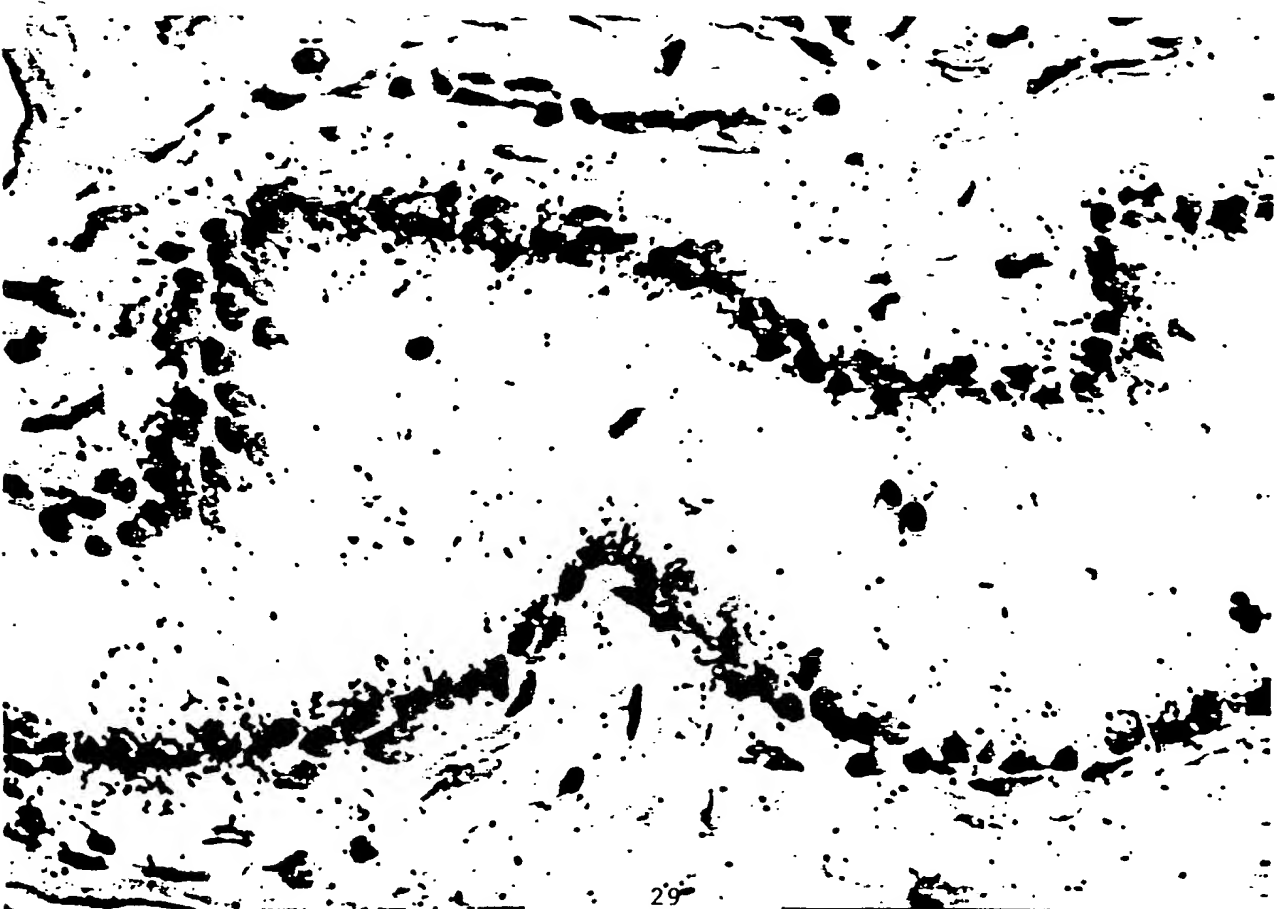


Fig. 10b

Mapping of hESX Activation Domain

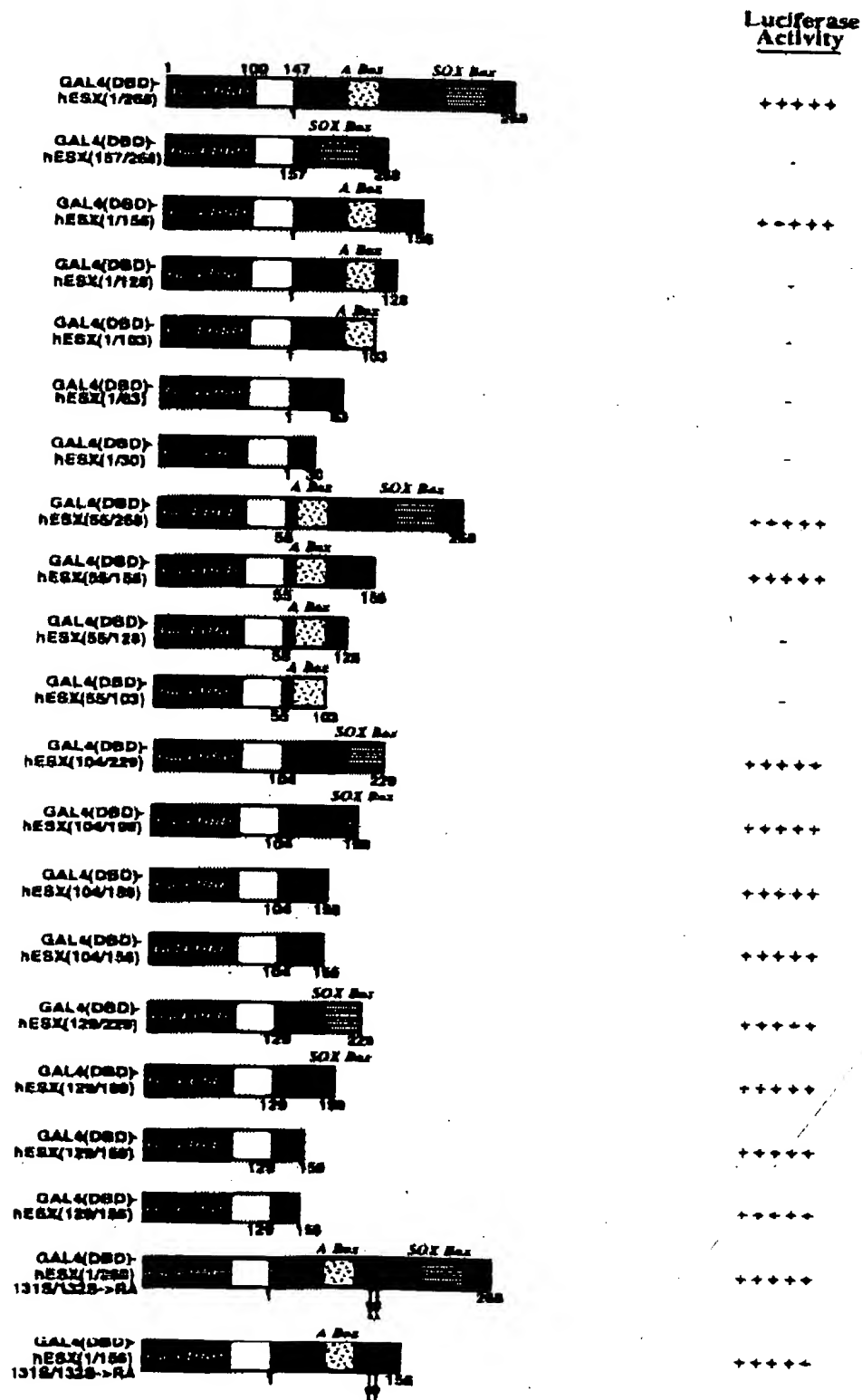


Fig. 11

08978217.11569